

WHAT IS CLAIMED IS:

1. A method of displaying files within a file system to a user in a semantic hierarchy, the method comprising the steps of:
 mapping the files into a semantic vector space;
 5 clustering the files within said space; and
 displaying the files in a hierarchical format based on the resulting clusters.
2. The method according to claim 1, wherein the step of clustering the files is performed as a background routine during the operation of a computer associated with said file system.
- 10 3. The method according to claim 2, wherein the step of clustering the files is performed in response to the creation of a new file within the file system.
4. The method according to claim 1, wherein said files are text documents and said mapping is conducted on the basis of a language model.
5. The method according to claim 4, wherein said mapping step comprises the
 15 steps of constructing a matrix which associates each word in the documents with a vector and associates each document with a vector.
6. The method of claim 5, further including the step of decomposing said matrix to define the words and documents as vectors in a continuous vector space.
7. The method of claim 5, wherein said clustering is performed by identifying
 20 documents whose vectors are within a threshold distance of one another.

8. The method of claim 7, further including the step of defining multiple threshold values and clustering said documents in accordance with said multiple threshold values to thereby establish plural levels of clusters.
- 5 9. The method of claim 5 further including the step of automatically labeling the clusters.
- 10 10. The method of claim 9 wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.
11. A graphical user interface configured to display files in a virtual file system with a semantic hierarchy.
12. The graphical user interface according to claim 11, wherein the semantic hierarchy is based on clustering of files based on semantic similarities.
13. The graphical user interface according to claim 12, wherein clustering of the files is initiated by user selection.
- 15 14. The graphical user interface according to claim 12, wherein clustering of the files is initiated upon creation of a new file in the file system.
15. The graphical user interface according to claim 12, wherein text files are clustered utilizing a language model and non-text files are clustered utilizing rule-based techniques.

16. The graphical user interface according to claim 15, wherein said language model comprises the LSA paradigm.
17. Computer readable media having stored therein computer executable code for analyzing files in a file system to determine similarities in data pertaining to their content, and displaying files in hierarchical format based on determined similarities between the files.
18. The computer-readable media of claim 17 wherein said files are text documents, and the similarities are based upon the word content of the files.
19. The computer-readable media of claim 18 wherein said similarities are determined in accordance with a language model, and the files are clustered in accordance with said model.
20. The computer-readable media of claim 19, wherein said language model comprises the LSA paradigm.
21. The computer-readable media of claim 19, wherein said computer-executable code performs the steps of constructing a matrix which associates each word in the documents with a vector and associates each document with vector.
22. The computer-readable media of claim 21, wherein said computer-executable code further performs step of decomposing said matrix to define the words and documents as vectors in a continuous vector space.

23. The computer-readable media of claim 22, wherein said computer-executable code performs clustering by identifying documents whose vectors are within a threshold distance of one another.
24. The computer-readable media of claim 23, wherein said computer-executable code further performs step of clustering said documents in accordance with multiple threshold values to thereby establish plural levels of clusters.
25. The computer-readable media of claim 19, wherein said computer-executable code performs step of automatically labeling the clusters.
26. The computer-readable media of claim 25, wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.
27. The computer readable media according to claim 16, wherein the computer executable code performs the following steps:
 - clustering text files within the file system using semantic similarities;
 - clustering non-text files within the files system using rule-based techniques;
 - labeling the resulting clusters; and
 - displaying the files in a hierarchical format based on the resulting clusters and labels.
28. A computer system, comprising:
 - a file system storing files;
 - a display device; and

a user interface which displays representations of files stored in said file system in the form of a semantic hierarchy that is based upon the content of said files.

29. The computer system of claim 28 further including a processor for
5 analyzing the content of files stored in said file system to map said files into a semantic vector space and cluster the files within said space, and wherein said user interface displays said files in accordance with said clustering.
30. The computer system of claim 29 wherein said files are text documents and said processor maps said files on the basis of a language model.
- 10 31. The computer system of claim 30 wherein said processor constructs a matrix which associates each word in the documents with a vector and associates each document with a vector.
32. The computer system of claim 31 wherein said processor further
15 decomposes said matrix to define the words and documents as vectors in a continuous vector space.
33. The computer system of claim 31, wherein said processor clusters the files by identifying documents whose vectors are within a threshold distance of one another.
- 20 34. The computer system of claim 33, wherein said processor clusters said files in accordance with multiple threshold values to thereby establish plural levels of clusters.

35. The computer system of claim 31, wherein said processor automatically labels the clusters.

36. The computer system of claim 35 wherein said processor labels the clusters by selecting representative words based on the closeness of their vectors to the
5 document vectors in a cluster.